

ABSTRACT OF THE DISCLOSURE

Conserved features of the sequence of a nucleic acid being sequenced are used to provide calibration markers for alignment of the data traces. Thus, in accordance with the invention, alignment of a plurality of data traces indicative of the positions of a plurality of nucleic acid base types in a target nucleic acid is achieved by selecting a set of three or more alignment points for each data trace. The alignment points are selected from among (1) a primer peak associated with unextended primer, (2) a full-length peak associated with full length product produced during a cyclic primer extension reaction or coupled amplification and sequencing reaction with one or two primers, and (3) internal peaks associated with internal bases that are highly conserved in the target nucleic acid. Each of the alignment points is associated with a reference position number reflecting the position of the alignment point with respect to the sequence as a whole. Sequencing position numbers are then assigned to each peak in each of the plurality of data traces in such a way as to maximize the number of times that the sequencing position number and the matching reference position number are assigned to a base of the same type. Finally, the data traces are aligned based on the assigned sequencing position numbers to provide aligned data traces for base-calling.